International Journal of
Physical Medicine & Rehabilitation

Dahl and Jørgensen, Int J Phys Med Rehabil 2014, 2:1
DOI: 10.4172/2329-9096.1000177

Research Article Open Access

Intra- and Inter-Rater Reliability of the Mini-Balance Evaluation Systems Test in Individuals with Stroke

Stine Susanne Haakonsen Dahl[1] and Lone Jørgensen[2,3]*

[1]Dahl Physiotherapy, Bodø, Norway
[2]Department of Health and Care Sciences and the "Tromsø Endocrine Research Group", University of Tromsø, Tromsø, Norway
[3]Department of Clinical Therapeutic Services, University Hospital of North Norway, Tromsø, Norway

Abstract

Objective: The aim of this study was to assess intra- and inter-rater reliability of the 'The Mini-Balance Evaluation Systems Test (Mini-BESTest)' in adults with stroke, based on video recordings of their test performances.

Methods: We included 24 adults with stroke classified at four different ambulatory levels, ranging from the ability to ambulate within the household only, to normal ambulation. Mini-BESTest performance of the participants were filmed and then scored by three raters twice, with four weeks between the sessions. None of the raters had used the test prior to the study, but attended a three-hour training session to become familiar with the test and scoring instructions just before study-start.

Relative reliability was investigated for by calculating intraclass correlation coefficients ($ICC_{1.1}$ and $ICC_{3.1}$). Absolute reliability was assessed by calculating within-subject standard deviation (s_w) and smallest detectable difference (SDD). For each individual item of the Mini-BESTest, Cohen's kappa (*k*) was calculated.

Results: The study showed that the Mini-BESTest had excellent intra-rater reliability ($ICC_{1.1}$=0.94−0.99 and $ICC_{3.1}$=0.97−0.99) and inter-rater reliability ($ICC_{1.1}$=0.97−0.99 and $ICC_{3.1}$=0.97−0.99). Kappa values for the individual items ranged between 0.21 and 1.00. The majority of items (intra-rater=88%, inter-rater=78%) showed very good or good agreement. The smallest detectable change at 95% confidence interval was ≤4 points for intra-rater assessments and ≤ 3 points for inter-rater assessments.

Conclusions: We conclude that the reliability of the Mini-BESTest based on video recordings of adults with stroke is excellent, even though the ratings are performed by novice assessors.

Keywords: Postural balance; Outcome assessment; Reliability; Stroke

Introduction

Individuals with stroke frequently have balance disorders that can lead to a reduced level of mobility and an increased risk of falling [1,2]. Consequently, a comprehensive assessment of balance is important for both diagnostic and therapeutic purposes. Good clinical balance measures are key features for evaluating balance functions, directing treatment and predicting outcome [3,4].

While many balance measures assess whether or not a balance problem exists and the severity of the disorder [1,5-8], these measures do not guide clinicians in directing treatment. The Balance Evaluation Systems Test (BESTest) is an extensive measure developed to assess and differentiate six systems underlying functional balance [9]. The reliability, validity and sensitivity have shown to be high in adults with a wide range of balance disorders [9-11]. The disadvantage of the test is that it takes approximately 45 minutes to complete, and therefore may not be feasible in clinical practice [7]. The Mini-BESTest is a shortened version of the BESTest focusing on dynamic balance. It takes between 10 and 15 minutes to administer [12]. Recent studies have reported very good psychometric properties of the test in a variety of neurological conditions influencing balance [11-16]. Only one study has specifically examined the intra-rater and inter-rater reliability of the Mini-BESTest in individuals with stroke, and both were found to be high [13]. The intra-rater reliability was evaluated by repeating the Mini-BESTest within 10 days by the same rater in 30 patients. However, because the period between sessions was short, there is a risk that the results had been influenced by the raters memory. Moreover, the variability of the scorings may have been due to differences in the performances of the patients and the instructions of the test. By the use of video the time between the first and second assessment can be expanded without risking that the variability of the scoring is due to these differences. The purpose of the present study was therefore to assess the intra- and inter-rater reliability of the Mini-BESTest in individuals with stroke using video recordings.

Methods

Setting

This study was conducted from September to December 2011 at the Physiotherapy Department, Bodø Rehabilitation Unit, Bodø, Norway.

Participants

Twenty-four people with stroke were recruited through health professionals working in stroke rehabilitation. Inclusion criteria were: ≥ 18 years of age and ability to walk at least six meters independently (cane allowed). Individuals that could not follow the test instructions or had impaired balance due to other diagnosis were not included. In order to enable assessment of all available scores on the Mini-BESTest, the participants were strategically selected based on their functional ambulation level as classified by the Functional Ambulation Classification of the Hospital of Sagunto (FACHS) [17,18]. The FACHS,

***Corresponding author:** Lone Jørgensen, Department of Health and Care Sciences, University of Tromsø, N-9037 Tromsø, Norway, Tel: +47 77646443; E-mail: lone.jorgensen@uit.no

Received November 15, 2013; **Accepted** January 16, 2014; **Published** January 20, 2014

Citation: Dahl SSH, Jørgensen L (2014) Intra- and Inter-Rater Reliability of the Mini-Balance Evaluation Systems Test in Individuals with Stroke. Int J Phys Med Rehabil 2: 177. doi:10.4172/2329-9096.1000177

developed for individuals with stroke, categorize the patients into 6 groups: level 0-1; inability to walk or requiring person assistance of 1 to walk, level 2; ability to ambulate within the household, level 3; ambulation in surroundings of the house and neighbourhood, level 4; independent community ambulation and level 5; normal ambulation [17]. The participants in the present study belonged to FACHS level 2 to 5, six persons at each level. One of the authors of this paper (SSHD), who was not one of the raters of the Mini-BESTest, assigned the FACHS scores.

Age, sex and information about the stroke (date of incident, type of stroke, brain localization and hemiparetic side) were registered for all participants.

The Regional Committee for Medical Health Research Ethics in Norway approved the study and the participants gave written informed consent.

Raters

Three raters, labelled A, B and C, were selected among physiotherapists working clinically with stroke patients. Rater A, B and C had worked as physiotherapists for 17, 16 and 1.5 years, respectively. All raters were blinded to the participants FACHS score. None of them were familiar with the Mini-BESTest prior to this study.

The Mini-BESTest

The Mini-BESTest [12] consists of 14 items, grouped into 4 systems of dynamic balance: "Anticipatory postural adjustments", "Postural responses", "Sensory orientation" and Balance during gait". All items are scored on an ordinal scale were 0 is severe, 1 is moderate and 2 is normal performance. For item 3 (stand on one leg) and item 6 (lateral compensatory stepping reactions), only the worse score is used when the total score is calculated. Thus the total score adds to a maximum of 28 points. The Mini-BESTest consists of standard instructions for patients and raters, and a description of what equipment to use [12]. In the present study a few modifications from the original Mini-BESTest were applied: For the compensatory stepping reaction tasks (item 4, 5 and 6) the participants were given two trials instead of one, and the best trial was scored. Five participants were unable to count backwards in three (when tested in sitting) and were given alternative tasks for the dual-task component of the timed "Up & Go" test (item 14); individuals with aphasia (n=2) were asked to walk with a cup of water [19], and the remainder three were asked to list girls names from A-Z in alphabetical order. All participants´ native language were Norwegian, and the instructions to the participants were therefore given in Norwegian. As no formal Norwegian translation of the Mini-BESTest was available when this study was conducted, instructions to the participants were translated from the original test (by author SSHD). For all other purposes the original text in English [12] was used.

Procedures

Prior to the main study, the three raters attended a three-hour training session to become familiar with the test and scoring instructions. The raters were given a copy of the Mini-BESTest and general information about the test. Each item of the test was demonstrated in the same room and with the same equipment as used for the study sample, and each of the score alternatives were discussed. The raters then watched the original BESTest training DVD and video clips of three adults performing the Mini-BESTest. Video clips were discussed to obtain a common understanding of the scores, and then scored independently by each rater. Training was provided by author SSHD, who had 10 years of working experience in Neurological Physiotherapy, and had attended a Mini-BESTest training course by one of the developers of the test (Fay Horak). Testing and filming procedures were piloted on one adult with a neurological disorder and two healthy adults.

All participants completed the Mini-BESTest in a quiet room in the Physiotherapy Department, using the same procedure and equipment. The participants wore shorts or equivalent, and flat or no shoes. One person wore an ankle and foot orthosis and two individuals used a cane for walking. Author SSHD, who was not one of the raters, instructed the participants in performing the test, while an assistant filmed their performance. The participants were informed that they were allowed to rest at any time. The Mini-BESTest took from 15 to 20 minutes to complete for each participant.

The sessions were recorded using a handheld-camera following a standardized procedure where the angle, height and distance of the camera were adapted from the procedures used in the original BESTest instructional-video. For items allowing two trials, both trials were recorded. The participants were filmed from when the instructions of the task were given, to the task was completed.

All three raters (A, B, C) scored the video clips of the twenty-four adults twice, with four weeks between the first (A1, B1, C1) and second (A2, B2, C2) rating sessions. Each rater assigned scores independently, but from the same video recordings, at the same time and in the same room. They were instructed not to discuss the scores with each other during or between the two sessions. For each new day of rating, the raters started by watching video clips of a healthy person performing the test. The participants were shown in random orders that differed from session one to session two. Individual items were shown in the same order as on the Mini-BESTest, and one item was scored before the next item was shown. When more than one trial was recorded, the raters were instructed to register the best score. The raters were allowed to watch each video clip several times, given that all raters watched all repetitions and scored after seeing the last one. Scores were registered on the Mini-BESTest standard assessment forms. A new form was distributed for each participant and the forms were unavailable for the raters after the assessment was completed.

Data analysis

Characteristics of the participants are presented as means, standard deviation (SD) and range for continuous variables or percentages for dichotomous variables. The correlation between the Mini-BESTest total score (mean value of all 6 ratings) and the FACHS score was examined using the Spearman rho.

The total scores of the Mini-BESTest were assessed for normality using the Kolmogorov-Smirnov test. As the test results were not statistically significant ($p>0.05$), parametric statistics could be used for analysis [20].

Relative reliability of the Mini-BESTest total score was assessed with Intraclass correlation coefficients (ICC) [20, 21]. ICCs were calculated to assess pairwise correlation between raters (A1-B1, A1-C1, B1-C1), within the group of raters (A1-B1-C1), and within raters (A1-A2, B1-B2, C1-C2). ICC (1.1) was used because the raters were strategically chosen [22]. As ICC (1.1) assumes all errors to be random measurement errors, ICC (3.1) was used in addition to ICC (1.1) enabling investigation for systematic errors. When ICC (1.1) equals ICC (3.1), no systematic errors are present [22]. The ICC ranges from 0 to 1, where 0 indicates no agreement and 1 perfect agreement. ICC values of 0.90 and above indicate very high agreement between measures [20].

Absolute reliability was investigated by calculations of within-subject standard deviations (s_w). S_w assess how a given sum score on

FACHS	Woman (n)	Men (n)	Age in years			Months after stroke (Mean)
			Mean	SD	Range	
level 2	4	2	64.2	13.7	(48-85)	44.5
level 3	2	4	61.2	9.9	(51-71)	57.2
level 4	1	5	58.5	20.7	(19-78)	105.3
level 5	4	2	54.5	10.2	(41-67)	9.7

FACHS: Functional Ambulation Classification of the Hospital of Sagunto.
FACHS, level 2; ambulation within the household, level 3; ambulation in surroundings of the house and neighbourhood, level 4; independent community ambulation, level 5; normal ambulation
SD: standard deviation.

Table 1: Characteristics of the participants (n=24) according to their ambulatory level assessed by FACHS.

Rater	Mini-BESTest total score		
	Mean	SD	Range
A1	13.8	6.1	1-27
A2	14.0	6.8	1-27
B1	13.4	6.9	0-27
B2	13.4	6.8	1-26
C1	13.3	6.7	0-25
C2	12.5	6.6	0-26

A1, B1, C1 refer to the first scores of the three raters. A2, B2, C2 refer to the second scores (four weeks later).
SD: standard deviation.

Table 2: Mini-BESTest total scores for the participants (n=24) according to each rater.

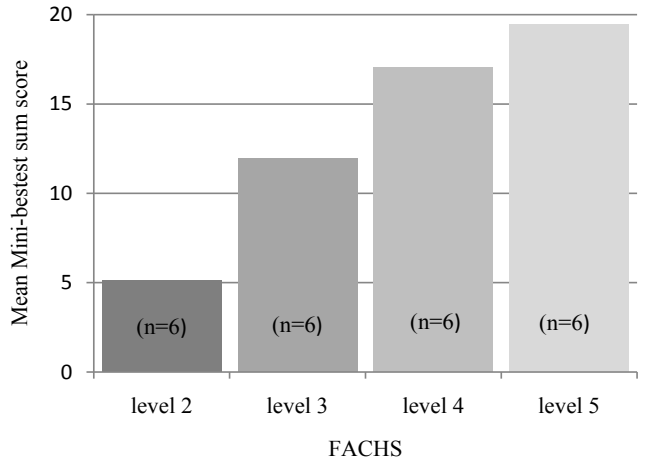

Figure 1: Mean Mini-BESTest sum scores for the 24 participants in relation to their Functional Ambulation Classification of the Hospital of Sagunto (FACHS) levels.
Level 2; ambulation within the household, level 3; ambulation in surroundings of the house and neighbourhood, level 4; independent community ambulation, level 5; normal ambulation

the Mini-BESTest is related to a "true" score for that person, and the variability in total scores with repeated observations, expressed in scores on the Mini-BESTest [20]. The data were checked for heteroscedasticity (when the measurement error depends on the size of the score value), and because no heteroscedasticity was found the calculations of s_w was justified. S_w was calculated using analysis of variance (ANOVA), where s_w equals the square root of the within-people residual mean square [23]. The difference between a participant's score assigned by one rater and the "true" score is expected to be <1.96 s_w for 95% of the observations. Thus the difference between two scores for the same participant is expected to be <$\sqrt{2}$×1.96 s_w for 95% of the pairwise observations [23]. This value is an estimate of the minimum change in score that is needed to be sure that the change is greater than the measurement error, and is referred to as the smallest detectable difference (SDD) [21].

Kappa (*k*) statistics was used to analyze degree of intra- and inter-rater agreement for each item of the Mini-BESTest [20]. The guidelines from Landis and Koch were used to interpret the results [20,24]. A *k* value of <0.20 is described as poor agreement, 0.21-0.40 fair, 0.41-0.60 moderate, 0.61-0.80 good and 0.81-1.00 is very good agreement [24]. As *kappa* can only be estimated when all score alternatives for an item are used, percentage agreement was calculated for items where some scores were not used. *Kappa* was the first choice as it corrects for chance agreement, while percentage of agreement does not [20].

Analyses were performed using IBM SPSS version 19.

Results

Participants

Characteristics of the participants according to their ambulatory levels are presented in Table 1. Of the 24 participants, 46% were woman, 17 had cerebral infarction, 6 had intra-cerebral haemorrhage and 1 had both infarction and haemorrhage. Ten had a right-sided hemiparesis and 8 had a left-sided hemiparesis and 6 had a lesion in the brainstem and/or cerebellum.

Mini-BESTest score

The total scores of the Mini-BESTest for all participants as given by each of the raters are shown in Table 2. The score of the participants ranged from 0 to 27, covering most of the available scores on the Mini-BESTest (ranging from 0 to 28). The Mini-BESTest total score correlated significantly with the FACHS score (Spearmans rho=.86, P<0.01). The relationship between the two tests is illustrated in Figure 1.

Reliability

Relative reliability of both intra- and inter-rater assessments was very high (ICC ≥ 0.94) (Table 3).

Within-subject standard deviation (s_w) and SDD for intra- and inter-rater reliability of the total score of the Mini-BESTest are also reported in Table 3. ANOVA calculation of within-people residual mean square was 1.014 for inter-rater analysis for all raters (A1-B1-C1), and from the equation $\sqrt{1.014}$ s_w was calculated to 1.0. The difference between a participant's total score and the "true" measurement value was then expected to be less than 2 points on the Mini-BESTest for 95% of the scores (± 1.96×1.0). The smallest detectable difference of the total Mini-BEST score between two measurements for the same participant

Raters	ICC(1.1)	95% CI	ICC(3.1)	95% CI	S_w	SDD
Intra-rater						
A1-A2	0.94	0.87-0.97	0.97	0.94-0.99	1.6	4.4
B1-B2	0.99	0.98-0.99	0.99	0.97-1.00	0.7	2.0
C1-C2	0.96	0.92-0.98	0.97	0.93-0.99	1.1	3.2
Inter-rater						
A1-B1	0.97	0.94-0.99	0.97	0.94-0.99	1.0	2.9
A1-C1	0.97	0.93-0.99	0.97	0.93-0.99	1.1	3.2
B1-C1	0.99	0.97-0.99	0.99	0.97-0.99	0.8	2.3
A1-B1-C1	0.98	0.95-0.99	0.98	0.95-0.99	1.0	2.8

A1, B1, C1 refer to the first scores of the three raters. A2, B2, C2 refer to the second scores (four weeks later).
ICC: intraclass correlation coefficient. CI: confidence interval. Sw: within subject standard deviation. SDD: smallest detectable difference for 95% of pairs of observations.

Table 3: Intra- and inter-rater reliability of the total score of the Mini-BESTest for the participants (n=24).

Item	Rater A1-A2 *k*	Rater B1-B2 *k*	Rater C1-C2 *k*	Rater A1-B1 *k*	Rater A1-C1 *k*	Rater B1-C1 *k*
Anticipatory postural adjustments						
1. Sit to stand	0.84	1.00	0.84	0.84	1.00	0.84
2. Rise to toes	0.74	1.00	0.80	0.60	0.74	0.72
3. Stand on one leg						
Left	0.87	1.00	0.94	0.87	0.80	0.94
Right	0.93	0.93	0.80	1.00	0.87	0.87
Postural responses						
4. Compensatory stepping correction - forward	0.55	0.85	0.84	0.49	0.56	0.77
5. Compensatory stepping correction - backward	0.68	0.68	0.87	0.87	0.74	0.87
Compensatory stepping correction - lateral						
Left	0.53	0.59	90%	0.67	71%	50%
Right	0.73	0.87	0.80	0.74	0.68	0.81
Sensory orientation						
7. Eyes open, firm surface	100%	0.81	95%	0.81	96%	98%
8. Eyes closed, foam surface	0.90	0.90	0.82	1.00	0.91	0.91
9. Incline – eyes closed	1.00	1.00	0.77	0.91	0.84	0.92
Balance during gait						
10. Change in gait speed	95%	0.75	0.70	95%	95%	0.54
11. Walk with head turns - horizontal	0.47	0.62	0.61	0.44	0.21	0.50
12. Walk with pivot turn	0.80	0.73	0.52	0.40	0.55	0.53
13. Step over obstacles	0.93	0.87	0.79	0.73	0.79	0.80
14. Timed up and go with dual task	0.87	0.80	0.86	0.73	0.73	0.86

Intra-rater (A1-A2, B1-B2, C1-C2) and inter-rater (A1-B1, A1-C1, B1-C1) agreement expressed in *k* (*kappa)* or % (percentage agreement).

Table 4: Intra- and inter-rater reliability of each item on the Mini-BESTest for the participants (n=24).
A1, B1, C1 refer to the first scores of the three raters. A2, B2, C2 refer to the second scores (four weeks later).

ranged between 2.0 and 4.4 points (95% CI) when scored by the same rater, and 2.3 and 3.2 points (95% CI) when scored by different raters.

Table 4 shows *Kappa* values (*k*) or percentages of agreement for each individual item. Overall, *k* ranged from 0.21 to 1.00. For intra-rater assessments 88% of all items showed good or very good agreement *(k* ≥ 0.61) and 12% moderate agreement (*k*=0.47-0.59). For inter-rater assessments 78% of all items showed good or very good agreement (*k* ≥ 0.67), 17% moderate (*k*=0.44-0.56) and 5% poor agreement (*k* ≤ 0.40). For intra-rater assessments, items 1 (sit to stand), 2 (rise to toes), 3 (stand on one leg, right) and 9 (incline, eyes closed) had perfect agreement (*k*=1), while item 11 (walk with head turn) showed the lowest pairwise agreement (*k*=0.47). For inter-rater assessments, item 1 (sit to stand), 3 (stand on one leg) and 8 (eyes closed, foam) showed perfect agreement (*k*=1), and item 11 (walk with head turns) had the lowest value (*k*=0.21). For ten items all score options were not used and therefore percentage agreements were calculated instead. The percentage of agreement ranged from 50% to 100% (mean 89%).

Discussion

Summary of results

This study showed, for both intra- and inter-rater assessments, very high relative and absolute reliability of the Mini-BESTest total score. The majority of the individual items had very good or good agreement, some moderate, and few had fair agreement.

Discussion of the results

The relative reliability was very high (ICC ≥ 0.94), meaning that the participants maintained their position in the group almost perfectly with repeated measurements [20]. Other studies examining the reliability of the Mini-BESTest have similar findings; in adults with stroke assessed in a live situation (intra-rater ICC=0.96, inter-rater ICC=0.97) [13], in adults with Parkinson's disease (inter-rater ICC=0.91) [11] and in adults with various balance disorders (inter-rater ICC=0.98, intra-rater ICC=0.96) [14]. However, one must have in mind that since the ICC value is higher if the sample encompasses a wide range of scores compared to a limited range, caution should be made when comparing the results from different studies [20].

The high ICC values should be considered together with the results of absolute reliability (reported below), as no single statistical analysis provides sufficient information on reliability on its own [20,21].

The low s_w for all raters (range 0.8–1.1 points) indicates a low measurement error of the Mini-BESTest. Calculations of SDD implies that a change in score of at least 4.4 points (intra-rater) and 3.2 points (inter-rater) can be interpreted as a real change (95% CI) when two measurements of the same participant are compared [23]. These results are in line with the results from a study of individuals with stroke (SDD=3.0 points) [13] and a study including individuals with a variety of balance disorders (SDD=3.5points) [15], but assessed in a live situation.

The individual items showing the highest agreement were item 1 (sit to stand), 2 (rise to toes), 3 (stand on one leg), 7 (eyes open, firm surface), 8 (eyes closed, foam) and 9 (incline, eyes closed). The scorings of these items are based on observations of tasks with only few components and/or time registered by a stopwatch. Scoring of such variables tends to show higher agreement than more complex tasks and tasks based solely on judging performance from observation [9]. This may also be the reason why item 4 (compensatory stepping correction–forward), 6 (compensatory stepping correction–lateral), 10 (change in gait speed), 11 (walk with head turns) and 12 (walk with pivot turns), showed the lowest agreements. With regard to balance during gait (item 11-14) the participants were mainly viewed from the front. Because movement strategies in individuals with stroke are complex, it is possible that the agreement had improved if the raters had been given the possibility to observe the participants from more than one plane of movement. As for the postural responses (item 4-6), it has been

reported that agreement is higher if the raters administer these items themselves [9].

The agreement between each pair of raters hardly differed. This suggests that the results of the Mini-BESTest are independent of years of working experience if only the physiotherapists have received training in how to use the test.

Discussion of the methods

The comprehensive analysis of rater reliability used in this study adheres to current recommendations for evaluation of clinical balance measures [20,21,25]. Moreover, the highly standardized procedures are considered a methodical strength.

Although there are no standard criteria regarding the time interval between assessments, enough time has to elapse to minimize the influence of the rater's memory when intra-rater reliability is examined. During this time there is a great probability that the patients have changed. By the use of video we ensure that the variability of the scoring is not due to differences in the patient's performances. In our study there were four weeks between the two rating sessions. The almost identical ICC (1.1) and ICC (3.1) implies that there were no systematic shift in data during this period of time [22], and indicates that neither a learning effect had taken place or that the raters required new training between session [20]. Scoring quality of movement is considered demanding and complex, but as video-recordings offers the possibility to view the same performance several times the accuracy of the scoring may improve. A limitation of the study is that the judgment of a therapist watching a performance from video-recordings may not be a complete reflection of the judgment made in live situations, where the raters may observe clinical parameters that are not visible on video tape. Moreover, when a test is rated in "real-time", the complexity of the whole test condition is evaluated, including the relation between the assessor and the subject, and the difficulty instructing and scoring the test simultaneously. Thus, each methodological approach has its advantages and disadvantages. The modifications of the Mini-BESTest items: allowing 2 trials for the postural responses (item 4-6) were considered appropriate. Similar modifications were also applied in a study assessing reliability of the Mini-BESTest in Parkinson's disease [11]. With regard to the dual-task modification (item 14) one could argue that the influence of holding a glass of water induces different restrictions to the degrees of freedom of the body and to the gait characteristics, in comparison to a cognitive dual-task**.** Since this modification was applied to two persons only, the effect is likely small. It is, however, important to point out that the cognitive dual-task in the Mini-BESTest is problematic for individuals with speech problems. The modifications we did may slightly have increased the time used to complete the Mini-BESTest, 15-20 minutes in the present study versus 10-15 minutes reported by others [12]. However, the simultaneous filming procedure also contributed to the increased length of time.

Strategic sampling of participants and raters may be a limitation to the external validity of this study [20]. On the other hand, the wide group of participants and raters add to the generalisability of the results. Furthermore, by selecting participants at different FACHS levels almost the whole range of the Mini-BESTest score (27 of a total of 28) were assessed. While the sample included a wide range of individuals with stroke in terms of demographics, ambulatory levels and the Mini-BESTest scores, individuals with major cognitive impairments were not included and the results can therefore not be generalized to this group of patients. Moreover, since all participants were Norwegians, and no formal Norwegian translation of the Mini-BESTest was available when this study was conducted, the instructions of the test were directly translated from English to Norweigan by the author. Considering the translatory variable, possible cultural discrepancy should be taking into account.

In conclusion, the Mini-BESTest seems to be a reliable measure for testing balance in adults with stroke, even though the ratings are performed by novice assessors.

Acknowledgement

This study was funded by The Rehabilitation Unit, Bodø, Norway and The Norwegian Fund for Post-Graduate Training in Physiotherapy.

References

1. Weerdesteyn V, de Niet M, van Duijnhoven HJ, Geurts AC (2008) Falls in individuals with stroke. J Rehabil Res Dev 45: 1195-1213.
2. de Oliveira CB, de Medeiros IR, Frota NA, Greters ME, Conforto AB (2008) Balance control in hemiparetic stroke patients: main tools for evaluation. J Rehabil Res Dev 45: 1215-1226.
3. Finch E, Brooks D, Stratford PW, Mayo NE (2002) Physical rehabilitation outcome measures; a guide to en hanced clinical decision making. (2ndedn) Ontario: Canadian Physiotherapy Association, B.C. Decker Incorporated publishers, Canada.
4. Shumway-Cook A, Wollacott M. Motor Control (2007) Translating Research into Clinical Practice. (3rd Edn.) Baltimore: Lippincott Williams & Wilkins, USA.
5. Podsiadlo D, Richardson S (1991) The timed "Up & Go": a test of basic functional mobility for frail elderly persons. J Am Geriatr Soc 39: 142-148.
6. Berg KO, Wood-Dauphinee SL, Williams JI, Maki B (1992) Measuring balance in the elderly: validation of an instrument. Can J Public Health 83 Suppl 2: S7-11.
7. Mancini M, Horak FB (2010) The relevance of clinical balance assessment tools to differentiate balance deficits. Eur J Phys Rehabil Med 46: 239-248.
8. Pollock C, Eng J, Garland S (2011) Clinical measurement of walking balance in people post stroke: a systematic review. Clin Rehabil 25: 693-708.
9. Horak FB, Wrisley DM, Frank J (2009) The Balance Evaluation Systems Test (BESTest) to differentiate balance deficits. Phys Ther 89: 484-498.
10. Leddy AL, Crowner BE, Earhart GM (2011) Functional gait assessment and balance evaluation system test: reliability, validity, sensitivity, and specificity for identifying individuals with Parkinson disease who fall. Phys Ther 91: 102-113.
11. Leddy AL, Crowner BE, Earhart GM (2011) Utility of the Mini-BESTest, BESTest, and BESTest sections for balance assessments in individuals with Parkinson disease. J Neurol Phys Ther 35: 90-97.
12. Franchignoni F, Horak F, Godi M, Nardone A, Giordano A (2010) Using psychometric techniques to improve the Balance Evaluation Systems Test: the mini-BESTest. J Rehabil Med 42: 323-331.
13. Tsang CS, Liao LR, Chung RC, Pang MY (2013) Psychometric properties of the Mini-Balance Evaluation Systems Test (Mini-BESTest) in community-dwelling individuals with chronic stroke. Phys Ther 93: 1102-1115.
14. Bergström M, Lenholm E, Franzén E (2012) Translation and validation of the Swedish version of the mini-BESTest in subjects with Parkinson's disease or stroke: a pilot study. Physiother Theory Pract 28: 509-514.
15. Godi M, Franchignoni F, Caligari M, Giordano A, Turcato AM, et al. (2013) Comparison of reliability, validity, and responsiveness of the mini-BESTest and Berg Balance Scale in patients with balance disorders. Phys Ther 93: 158-167.
16. King LA, Priest KC, Salarian A, Pierce D, Horak FB (2012) Comparing the Mini-BESTest with the Berg Balance Scale to Evaluate Balance Disorders in Parkinson's Disease. Parkinsons Dis 2012: 375419.
17. Viosca E, Martínez JL, Almagro PL, Gracia A, González C (2005) Proposal and validation of a new functional ambulation classification scale for clinical use. Arch Phys Med Rehabil 86: 1234-1238.
18. Viosca E, Lafuente R, Martínez JL, Almagro PL, Gracia A, et al. (2005) Walking recovery after an acute stroke: assessment with a new functional classification and the Barthel Index. Arch Phys Med Rehabil 86: 1239-1244.
19. Shumway-Cook A, Brauer S, Woollacott M (2000) Predicting the probability for falls in community-dwelling older adults using the Timed Up & Go Test. Phys Ther 80: 896-903.

20. Carter RE, Lubinsky J, Domholdt E (2011) Rehabilitation Research: Principles and Applications. (4th Edn.) St. Louis: Elsevier Saunders, USA.

21. Moe-Nilssen R, Nordin E, Lundin-Olsson L; Work Package 3 of European Community Research Network Prevention of Falls Network Europe (2007) Criteria for evaluation of measurement properties of clinical balance measures for use in fall prevention studies. J Eval Clin Pract 14: 236-240.

22. Shrout PE, Fleiss JL (1979) Intraclass correlations: uses in assessing rater reliability. Psychol Bull 86: 420-428.

23. Bland JM, Altmann DG (1996) Measurement error. BMJ 313:744.

24. Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. Biometrics 33: 159-174.

25. Küçükdeveci AA, Tennant A, Grimby G, Franchignoni F (2011) Strategies for assessment and outcome measurement in physical and rehabilitation medicine: an educational review. J Rehabil Med 43: 661-672.